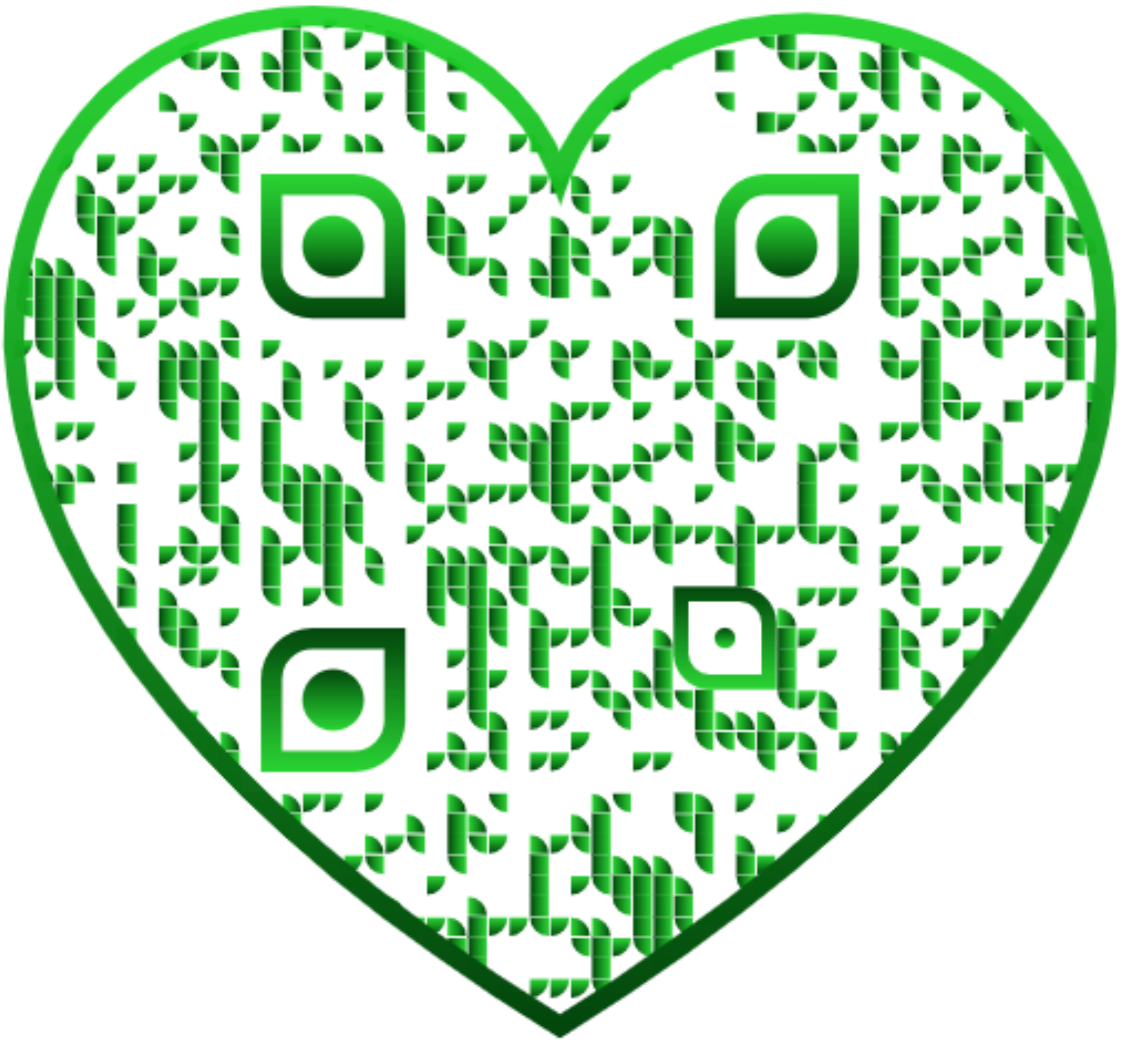
Chapter 2
Methods of Data Collection and Presentation

At the end of this chapter, the students will be able to:

1. Describe the sources of data
2. Identify the different methods of data collection
3. Identify different methods of data organization and Presentation methods (tabular and pictorial)
4. Determine the criterion for the selection of a method to organize and present data



Data Collection Methods

- ✓ Before any statistical work, data must be collected.
- ✓ Depending on the type of variable and the objective of the study different data collection methods can be employed.
- ✓ Data collection techniques allow us to systematically collect data about our objects of study.
- ✓ If data are collected haphazardly, it will be difficult to answer our research questions in a conclusive way.
- ✓ Therefore, it should be systematic/scientific.

Data Collection Methods

- ✓ Using available information
- ✓ Observation
- ✓ Interviewing (face-to-face)
- ✓ Self-administered written questionnaires
- ✓ Postal or mail method and telephone interviews
- ✓ Focus group discussions (FGD)

1) Using available information

- Examples include:
  1. Official publications of Central Statistical Authority
  2. Publication of Ministry of Health and Other Ministries
  3. News Papers and Journals.
  4. International Publications like Publications by WHO, World Bank, UNICEF.
  5. Records of hospitals or any Health Institutions.

Using available information…

- During the use of data from documents, though they are less time consuming and relatively have low cost, care should be taken on the **quality** and **completeness** of the data.
- Advantage of available data – collection is inexpensive
- Disadvantage of existing data:
  - It is some times difficult to gain access to records or reports
  - Data may not always be complete and precise enough, or too disorganized

2) Observation

- **Observation** is a technique that involves systematically selecting, watching and recoding behaviors of people and characteristics of living beings, objects or phenomena.
- They can also check on the information collected through interviews especially on sensitive topics such as drug use
- Observations can also be made on objects
- E.g the presence or absence of latrines and the state of cleanliness

Observation ...cont'd

- Observation of human behavior can be undertaken in different ways:

A) Participant observation – observer takes part in the situation he or she observes

- Eg. A doctor hospitalized with a broken hip, who now observes hospital procedures 'from within'

B) Non-participant observation – observer watches the situation, openly or concealed, but does not participate

- Eg. Mystery clients trying to obtain antibiotics without medical prescription

Observation ...cont'd

- ✓ **Advantages**: Gives relatively more accurate data on behavior and activities.
- ✓ **Disadvantages**: Investigators or observer's own biases, prejudice, desires, and etc.
- ✓ Needs **more resources** and skilled human power during the use of high level machines.

3) Interviewing

- It involves oral questioning of respondents, either individually or as a group
- Answers can be recorded by:
  - Writing them down
  - Tape-recording
  - Combination of them
- Interviews can be conducted with varying degree of flexibility (high degree of flexibility Vs low degree of flexibility)

Interviewing cont.

A) High degree of flexibility:

- Usually used when:
  - Studying sensitive issues or
  - When the researcher has little understanding of the problem
- Is frequently applied in exploratory studies
- When studying sensitive issues (e.g teenage pregnancy) the investigator may use a list of topics rather than fixed questions

Interviewing cont.

B) Low degree of flexibility

- Useful when:
  - Researcher is relatively knowledgeable about expected answers or
  - When the number of respondents being interviewed is relatively large
- Questionnaires may be used with a fixed list of questions in a standard sequence, which have mainly fixed answers

4) Administering written questionnaires (self-administered questionnaire)

- Written questions are presented that are to be answered by the respondents in written form.

Administering written questionnaires cont…

- ✓ Self-administered questionnaires is **simpler** and **cheaper**;
- ✓ Such questionnaires can be administered to many persons simultaneously
- ✓ Unlike interviews, can be sent by post or e-mail.
- ✓ But, they demand a certain level of education and skill on the part of the respondents

Administering written questionnaires cont…

- A written questionnaire can be administered in different ways, such as by:
  - Sending questionnaires by mail
  - Gathering all or part of the respondents in one place at one time, giving oral or written instructions, and letting them fill out the questionnaires
  - Hand-delivering questionnaires to respondents and collecting them later

Mailed questionnaire

- ✓ The **investigator prepares** a questionnaire containing a number of questions pertaining the field of inquiry.
- ✓ The questionnaires are **sent by post** to the informants together with a polite covering letter explaining the detail,
- ✓ The aims and objectives of collecting the information, and requesting the respondents to cooperate by furnishing the correct replies and returning the questionnaire properly filled in.

Mailed ...cont'd

- ✓ In order to ensure quick response, the **return postage expenses** are usually **borne by the investigator**.
- ✓ The main **problems with postal questionnaire** are;
  - ✓ Response rates tend to be **relatively low**, and
  - ✓ There may be under representation of less literate subjects.

What is Questionnaire?

- ✓ A questionnaire (or form) is a group or sequence of questions designed to obtain information on a subject from a respondent.
- ✓ Questionnaires include forms used for sample surveys and censuses as well as administrative forms.
- ✓ Questionnaires play a central role in the data collection process.
- ✓ They have a major impact on data quality since the questionnaire is the process whereby data are collected.

Importance of Questionnaire

- Is to provide a **standardized interview** across all subjects.
- They are always asked in **exactly the same way.**
- A well-designed questionnaire should:
  - Collect data efficiently with a **minimum number of errors and inconsistencies**;
  - Be **respondent friendly** and **interviewer friendly** (if interviewer-assisted);
  - Lead to an overall **reduction in the cost and time** associated with data collection.

Types of Questionnaire

1. Open-ended questions:

- ✓ **Permit free responses** that should be recorded in the **respondent's own words**.
- ✓ The respondent is not given any possible answers to choose from.
  - ❑ Such questions are useful to obtain information on:
  - ✓ Facts with which the researcher is **not very familiar**,
  - ✓ Opinions, attitudes, and suggestions of informants, or
  - ✓ **Sensitive issues**

Open ended ...cont'd

- ✓ Can you describe exactly what the traditional birth attendant did when your labor started?
- ✓ What do you think are the reasons for a high drop-out rate of village health committee members?
- ✓ What would you do if you noticed that your daughter had a relationship with a teacher?

2. Closed ended questions

- ✓ Closed questions offer a list of possible options or answers from which the respondents must choose.
- ✓ When designing closed questions one should try to:
  - ✓ Offer a list of options that are exhaustive and mutually exclusive
  - ✓ Keep the number of options as few as possible.
- ✓ Closed questions are useful if the range of possible responses is known.
- ✓ For example: “What is your marital status?
  1. Single
  2. Married/living together
  3. Separated/divorced/widowed

Closed ... Cont'd

- ✓ Closed questions may also be used if one is only interested in certain aspects of an issue and does not want to waste the time of the respondent and interviewer by obtaining more information than one needs.
- ✓ "Did you eat any of the following foods yesterday? (Circle yes or no for each set of items)
- ✓ Peas, bean, lentils Yes No
- ✓ Fish or meat Yes No
- ✓ Eggs Yes No
- ✓ Milk or Cheese Yes No

Steps in questionnaire design

- ✓ In the first draft we should concentrate on the **content**.
- ✓ In the second, we should look critically at the **formulation** and **sequencing** of the questions.
- ✓ Then we should **examine** the format of the questionnaire.
- ✓ Finally, we should do a **test-run** to check whether the questionnaire gives us the information we require and whether both the respondents and we feel at ease with it.

Steps ..cont'd

- ✓ **STEP1**: CONTENT: Take your objectives and variables as your starting point.
- ✓ **STEP 2**: FORMULATING QUESTIONS: Formulate one or more questions that will provide the information needed for each variable. Here consider the following points
- ✓ Check whether each question measures one thing at a time.
- ✓ For example, the question, "How large an interval would you and your husband prefer between two successive births?"
- ✓ Would better be divided into two questions because husband and wife may have different opinions on the preferred interval.

Steps ..cont'd

- ✓ **Avoid leading questions**: A question is leading if it suggests a certain answer.
- ✓ For example, the question, "Do you agree that the district health team should visit each health center monthly?" hardly leaves room for "no" or for other options.
- ✓ Better would be: "Do you thing that district health teams should visit each health center? If yes, how often?"
- ✓ **STEP 3**: SEQUENCING OF QUESTIONS: Design your interview schedule or questionnaire to be "**consumer friendly."**
- ✓ Sequence the questions in logical order; background, context
- ✓ Use simple language

Steps ..cont'd

- ✓ **STEP 4**: **Formatting** the Questionnaire
- ✓ Your questionnaire should not only be consumer but also **user friendly**!
- ✓ **STEP 5**: **Translation**: If interview will be conducted in one or more local languages, the questionnaire has to be translated to standardize the way questions will be asked.
- ✓ After having it translated you should have it retranslated into the original language. You can then compare the two versions for differences and make a decision concerning the final phrasing of difficult concepts.

G. Assignment-1

1. Assessment of the prevalence and risk factor of Anemia?
2. TB
3. Diarrial disease
4. Malaria
5. HIV/AIDS
6. Pneumonia
7. DM
8. Asthma
9. Utilization of Modern Contraceptives
10. Coverage and utilization of latrine
11. Assessment of Indoor Air quality

1) Write a title and objective,
2) List dependent and independent variable
3) Prepare questionnaire in English and translate to Amharic language

5. Focus Group Discussions

- A focus group is a discussion of a selected topic by participants who are chosen from the population of interest.
- Focus groups provide the opportunity to consult with members of the target population, data users, or interviewers so that their points of view can be brought into the questionnaire design process.
- In the early stages of questionnaire development, focus groups can be used to clarify survey objectives and data requirements, to identify main research issues, definitions, and concepts.
- Focus groups are also useful in **testing** questionnaires. They are used to evaluate a respondent's understanding of the language and wording used in questions and instructions, and to evaluate alternative question wordings and formats.

Focus Groups… cont'd

- ✓ A focus group is led by a **moderator** who is knowledgeable about group interviewing techniques and the purpose of the discussion.
- ✓ The size of each group is usually between 6 and 12 persons, with the optimum size being 7 to 9 persons.
- ✓ A focus group usually **lasts** for about **two hours**.
- ✓ Focus groups are **audio-recorded** (and sometimes video-recorded) and can be viewed by observers in an adjoining room behind a one-way mirror.

Choosing a Method of Data Collection

- ✓ Decision-makers need information that is relevant, timely, accurate and usable.
- ✓ The cost of obtaining, processing and analyzing these data is high.
- ✓ The challenge is to find ways, which lead to information that is cost-effective, relevant, timely and important for immediate use.
- ✓ Some methods pay attention to timeliness and reduction in cost.
- ✓ Others pay attention to accuracy and the strength of the method in using scientific approaches.
- ✓ Generally, the **choice of methods** of data collection is largely based on the **accuracy** of the information they yield.

Choosing ... Cont'd

- ✓ In this context, '**accuracy**' refers not only to correspondence between the information and objective reality - although this certainly enters into the **concept** - but also to the information's **relevance**.
- ✓ This issue is the extent to which the method will provide a precise measure of the variable the investigator wishes to study.

Choosing ... Cont'd

- ✓ Method of data collection is also based on practical considerations, such as:
  - ✓ The **need for personnel**, skills, equipment, etc. in relation to what is available and the urgency with which results are needed.
  - ✓ The **acceptability of the procedures** to the subjects - the absence of inconvenience & unpleasantness consequences.
  - ✓ The probability that the method will provide a **good coverage**, i.e. will supply the required information about all or almost all members of the population or sample.

Common problems in method of data collection

- ✓ Language barriers
- ✓ Lack of adequate time
- ✓ Expense
- ✓ Inadequately trained and experienced staff
- ✓ Violence of privacy
- ✓ Bias (person, season, diplomatic, professional)
- ✓ Cultural norms (e.g. which may preclude men interviewing women)

Methods of Data organization & presentation

- The data collected in a survey is called raw data.
- In most cases, useful information is not immediately evident from the mass of unsorted data.
- Collected data need to be organized in such a way as to **condense** the information they contain in a way that will show patterns of variation clearly.

Choosing graphs

Types of Data or Purpose	Appropriate Graph	Remark
Numerical (Interval or Ratio)	Histogram	Continuous data
	Frequency Polygon	
	C. Frequency Curve or OGIVE Curve	
	Scatter plot	
Categorical	Bar	
	Pie	
Trend	Line	

1) Frequency Distributions

- One of the most important means of summarizing the data from a single variable is to tabulate the frequency distribution of the variable.

  1) Ungrouped Frequency Distribution

  2) Grouped Frequency Distribution

- A frequency distribution simply tells how often a variable takes on each of its possible values.
- For example, Gender in this class (Nominal Var.)

Frequency Distribution of Gender of Nursing students, DBU, 2008

Gender	Frequency	Relative Frequency (Proportion)	Relative Frequency (Percentage)
Male	22	22/43 = 0.535	23/43*100 = 53.5%
Female	20	20/43 = 0.465	20/43*100 = 46.5%
Total	43		

Example (2): Top ten disease

- 200 patients were extracted from OPD registration book of certain health facility to prioritize the cases.
- Let us construct a frequency distribution table for this scenario

Frequency distribution of top ten diseases in a facility, 2015

Disease	Frequency	Relative Frequency	Cumulative Frequency proportion	Relative Frequency percentage
Diarrhea	**56**	56/200 = 0.28	56/200 = 0.28	**28%**
Pneumonia	**34**	34/200 = 0.17	90/200 = 0.45	**17%**
UTI	**25**	25/200 = 0.12	115/200=0.57	**12%**
STIs	**19**	19/200 = 0.09	134/200=0.67	**9%**
TB	**17**	17/200 = 0.08	151/200=0.75	**8%**
Asthma	**15**	15/200 = 0.07	166/200=0.83	**7%**
Dyspepsia	**12**	12/200 = 0.06	178/200=0.89	**6%**
Malaria	**10**	10/200 = 0.05	188/200=0.94	**5%**
Trachoma	**7**	7/200 = 0.035	195/200=0.97	**3.5%**
Teaniasis	**5**	5/200 = 0.025	200/200=1	**2.5%**
Total	**200**			

Grouped data frequency distribution

- ✓ **Classes:** Categories for grouping data.
- ✓ **Frequency (class frequency):** The number of data values in a class.
- ✓ **Relative frequency:** The proportion of the frequency of a class to the total number of pieces of data.
- ✓ **Frequency distribution:** A listing of classes and their frequencies.
- ✓ **Relative Frequency distribution:** A listing of classes and their relative frequencies.

Terms cont'd

- ✓ **Upper class limit:** The largest value that can go in a class.
- ✓ **Lower class limit:** The smallest value that can go in a class.
- ✓ **Class width:** The difference between the lower class limit of the given class and the lower class limit of the next higher class.
- ✓ **Class midpoint (class mark):** The midpoint of a class.

To build a Grouped Frequency Dist. Table

1. Find the range of the data: **Range** = Largest value – smallest value
2. Use the **range and try different class width to determine how many classes** you need to make frequency table or histogram.

Tabular presentation of data

Stubs → Horizontal			
Captions ↓ Vertical			
		Cell	

Statistical table can be:

1. **Simple or One Way Table**: used when the individual observations involve only to a **single** variable.
2. **Two Way Table**: shows **two characteristics** and is formed when either the caption or the stub is divided into two or more parts.
3. **Higher Order Table**: When it is desired to represent **three or more characteristics** in a single table.

Example (1): Simple

Table 1: Overall immunization status of children in Adami Tullu Woreda, Feb. 1995

Immunization Status	Number	Percentage
Not Immunized	75	35.7%
Partially Immunized	57	27.1%
Fully Immunized	78	37.2%
Total	210	100%

Source: Fikru T et al. EPI Coverage in Adami Tulu. Eth J Health Dev 1997;11(2): 109-113

Example (2): Two Way Table

Table 2: TT immunization by marital status of the women of childbearing age, Assendabo town, Jimma Zone, 1996

Marital Status	Immunization Status				Total
	Immunized		Non Immunized		
	No.	%	No.	%	
Single	58	24.7	177	75.3	235
Married	156	34.7	294	65.3	450
Divorced	10	35.7	18	64.3	28
Widowed	7	50	7	50	14
Total	231	31.8	496	68.2	727

Source: Mikael A. et al Tetanus Toxoid immunization coverage among women of child bearing age in Assendabo town; Bulletin of JIHS, 1996, 7(1): 13-20

Diagrammatic Representation

- It is Pictorial or graphic presentations of numerical data

Diagrammatic representation

- Frequency distributions can often be displayed effectively using graphical means such as the bar chart, pie chart or histogram.
- They follow the following principles
  - Have descriptive title (Include what, where and when)
  - Properly labeled axes
  - Indicate unit of measurement
  - Legends or key

Importance of Diagrammatic Representation

- They have greater attraction than mere figures.
- help in deriving information in less time and without any mental strain.
- They facilitate comparison.
- They have greater memorising value than mere figures.

Limitations of Diagrammatic Representation

- Diagrammatic representation is made use only **for purposes of comparison**.
  - It is not to be used when comparison is not possible or not necessary
- It only strengthens the textual exposition of a subject, and cannot serve as a complete substitute for statistical data
- It can give only an approximate idea
- They fail to bring to light small differences

2) Bar Graph

- ✓ **Data:** Nominal/Ordinal.
- ✓ **Horizontal** axis is the categories of the variable .
- ✓ **Vertical** axis is frequency or relative frequency
- ✓ **Height** of bar corresponds to frequency of category
- ✓ **Width** of the bars is constant.
- ✓ **Bars** separated by constant distance.

A. **Simple bar chart**: It is a one-dimensional diagram in which the bar represents the whole of the magnitude.

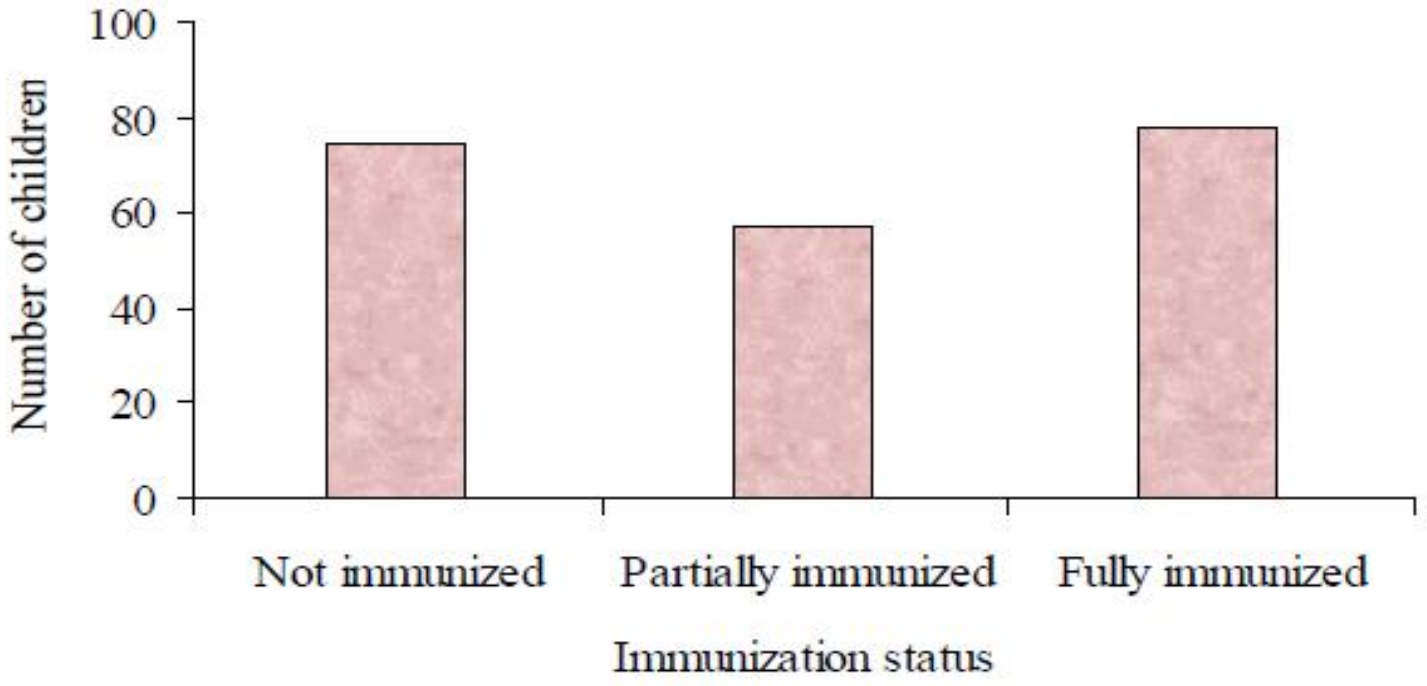


Fig. 1. Immunization status of children in Adami Tulu Woreda, Feb. 1995

B. Multiple bar chart: In this type of chart the component figures are shown as separate bars adjacent to each other.

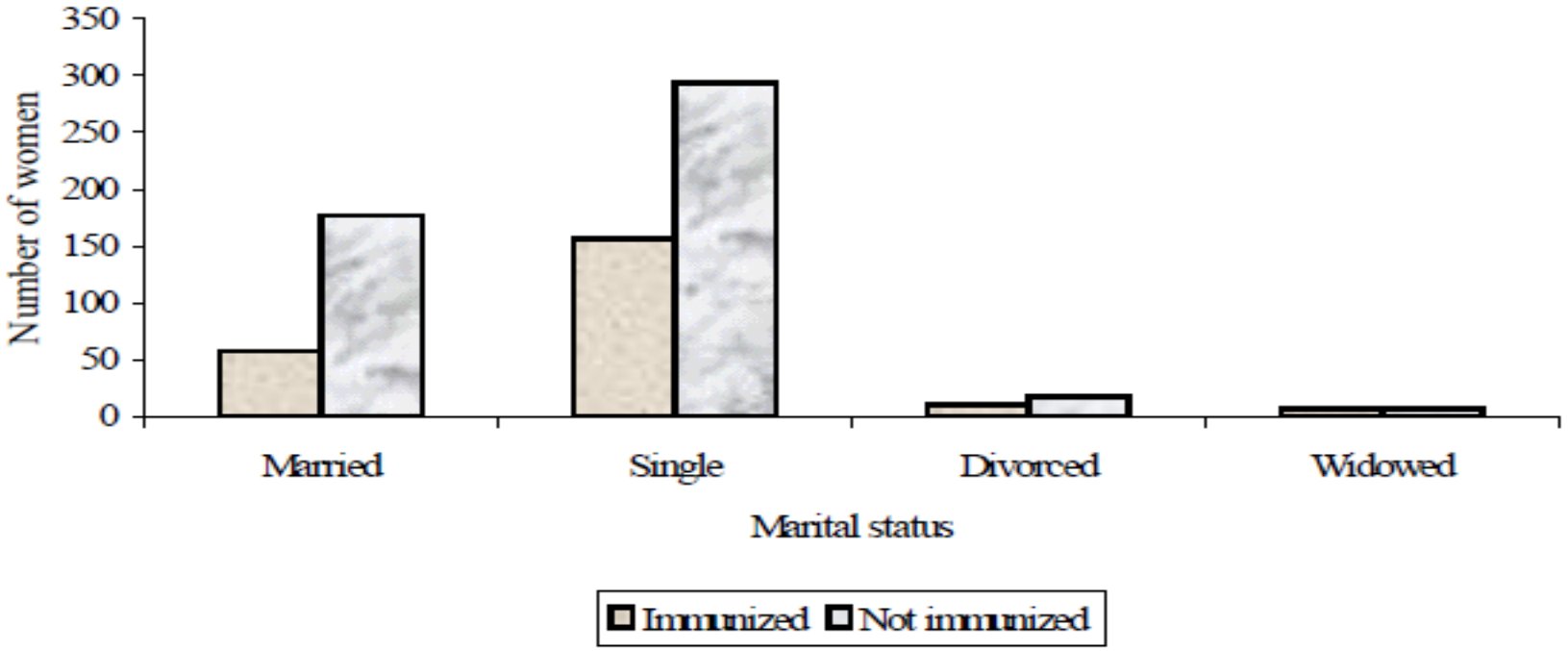


Fig. 2 TT Immunization status by marital status of women 15-49 years, Asendabo town, 1996

3) Pie chart

- **Pie-chart** (qualitative or quantitative discrete data): it is a circle divided into sectors so that the areas of the sectors are proportional to the frequencies.
- Pie charts depict the same information as do bar graphs, but in the shape of a circle or pie.
- The circle is divided into sectors, one for each category of the data.
- The area of each sector is Proportional to the frequency or relative frequency of each category of the variable.

Example of Pie Chart

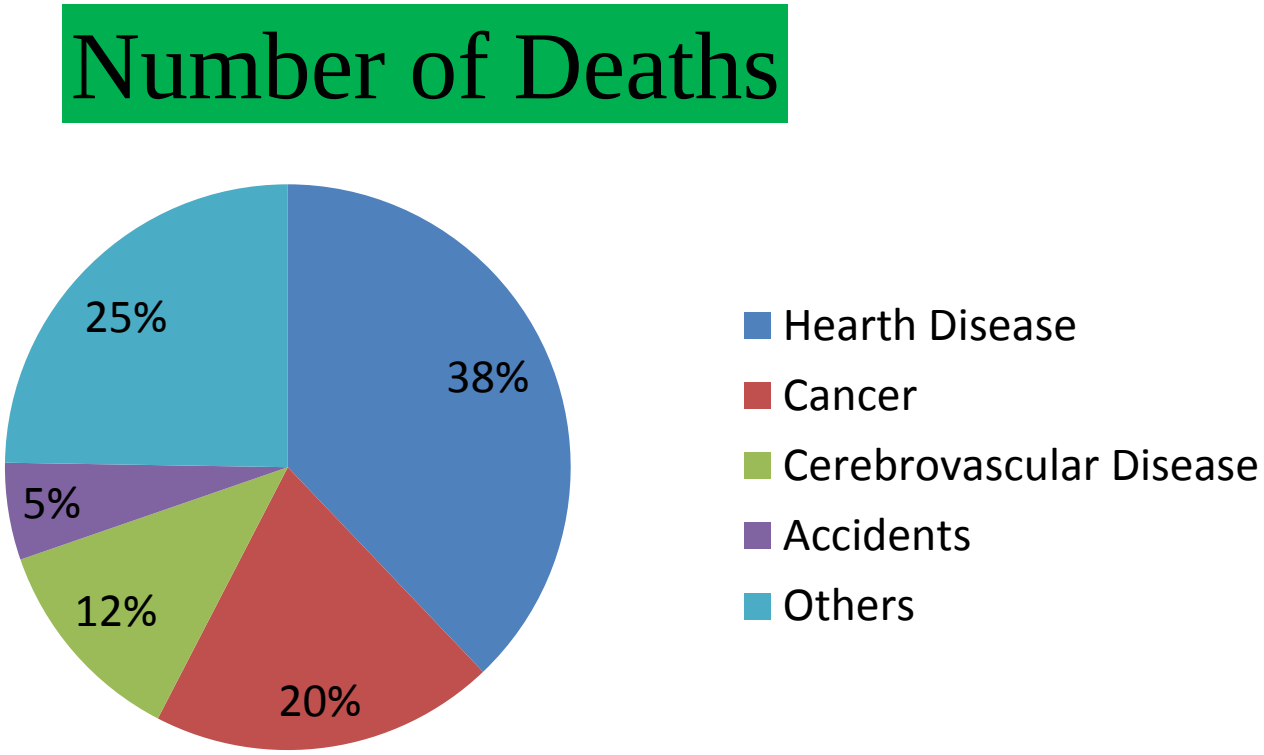


Figure 1: Cause of deaths in Minnesota, 1975

4) Histogram

- A graph of continuous **(Interval/Ratio)** variable in which

1) The vertical scale represents the frequency or relative frequency in each interval.

2) The horizontal scale represents the value of the variable marked at interval boundaries/or real limits.

✓ Therefore, touch Keep rectangles widths constant

Example (Checking asymmetry/or skewness of data)

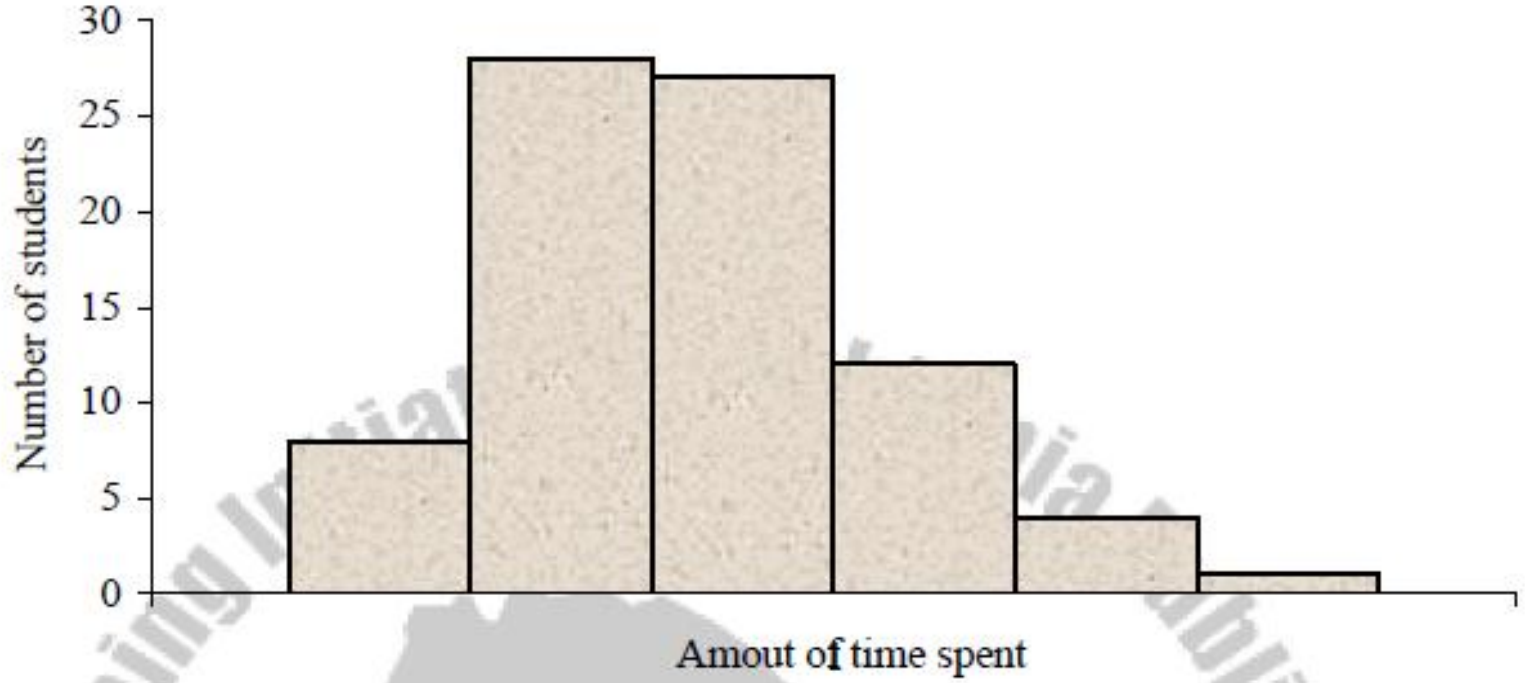


Figure 2: Amount of time University students spent on leisure activities, 2010

5) Frequency polygon

- If we join midpoints of the tops of the adjacent rectangles of histogram with line segments a **frequency polygon** is obtained.
- When the polygon is continued to the X-axis just out side the range of the lengths **the total area under the polygon will be equal to the total area under the histogram**

☐ It can be drawn with out erecting rectangles of histogram as follows:

1. scale should be marked in numerical values of the **midpoints of intervals.**
2. Erect ordinates on the midpoints of the interval - .
3. Join tops of ordinates & extend the connecting lines to the scale of sizes.

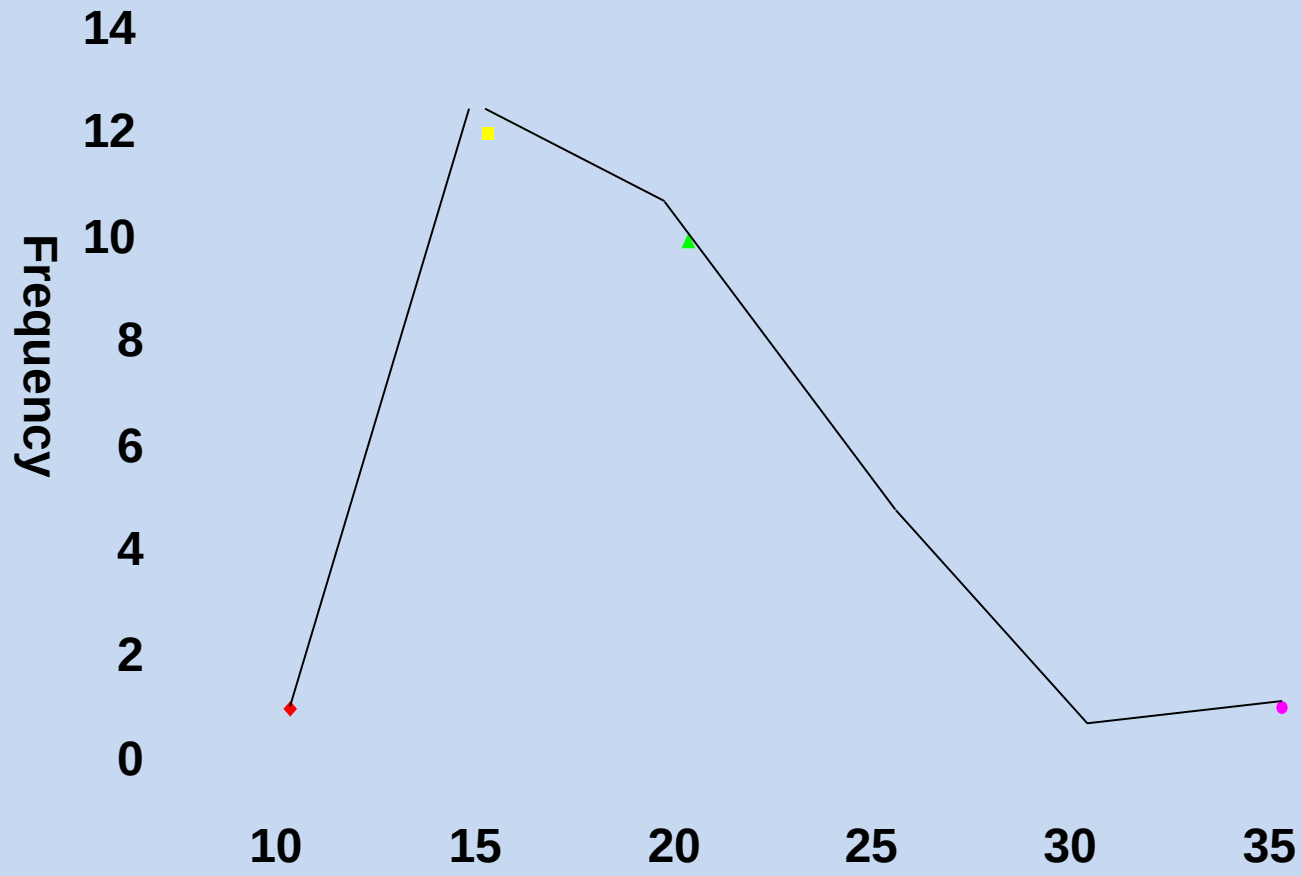
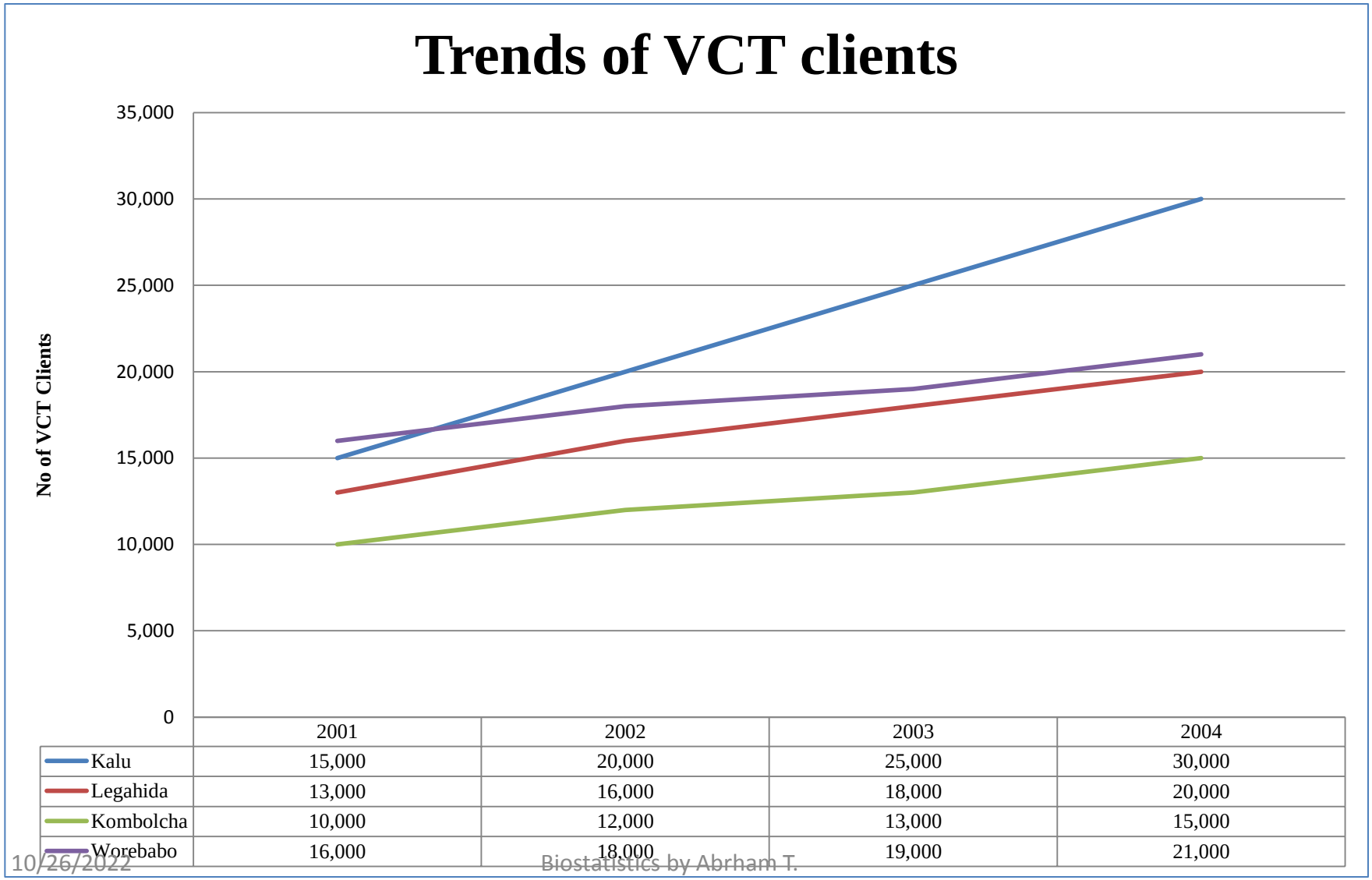


Fig 7: Frequency polygon curve on time spent for leisure activities by students

6) Line Graph

- ✓ Line graphs are good at showing overall **trends** of how a particular characteristic is performing over time.
- ✓ Suppose we are interested to visually compare VCT performances of facilities in four districts.
- ✓ We aggregated the number of clients visiting our facilities in the district and used single value per year to see their comparative performance over time.
- ✓ Data was available for four years and it was plotted using simple trend graph in excel as shown below.

Line graph … cont'd



	2001	2002	2003	2004
Kalu	15,000	20,000	25,000	30,000
Legahida	13,000	16,000	18,000	20,000
Kombolcha	10,000	12,000	13,000	15,000
Worebabo	16,000	18,000	19,000	21,000

Summary

1. What are the advantages and disadvantages of primary and secondary data?
2. What is the secrete behind data summarization?
3. List down the importance of pictorial representation of data.
4. Search the applications of Scatter Diagram, Histogram and line graph.
5. What are the five numerical measures used to draw Box and whisker Plot?

Reading assignment

- Q-Q plot
- P-P plot
- Steam and leaf plot
- Box and Whisker Plot
- Scatter plot

Stem-and-Leaf Plot…

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

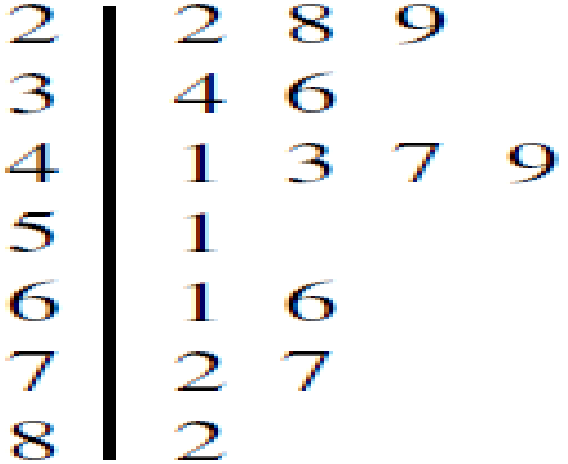
- Here, use the 10's digit for the stem unit:

- 12 is shown as → 1 | 2
- 35 is shown as → 3 | 5

Stem	Leaf
1	2
3	5

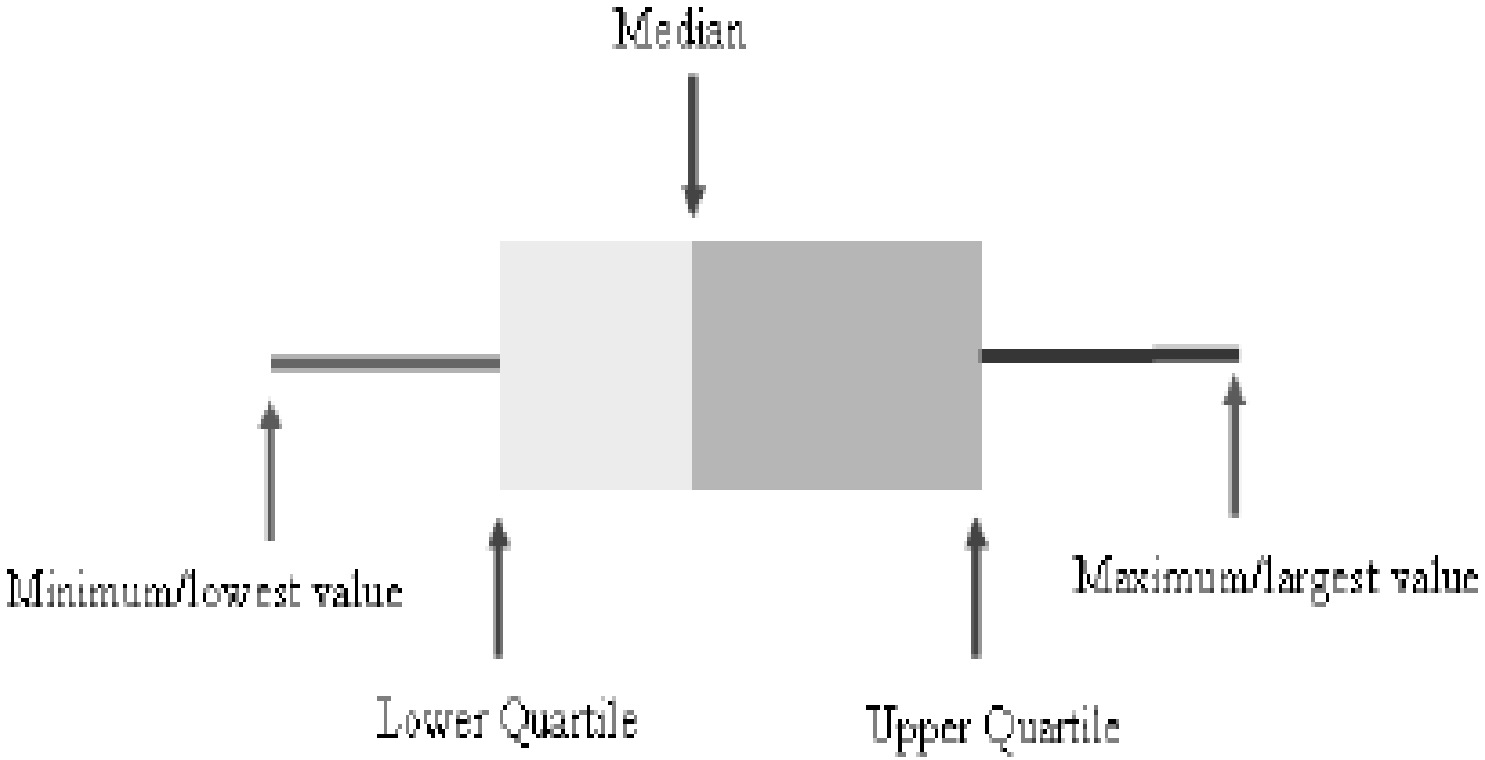
Example: Ordered Array

43, 28, 34, 61, 77, 82, 22, 47, 49, 51, 29, 36, 66, 72, 41

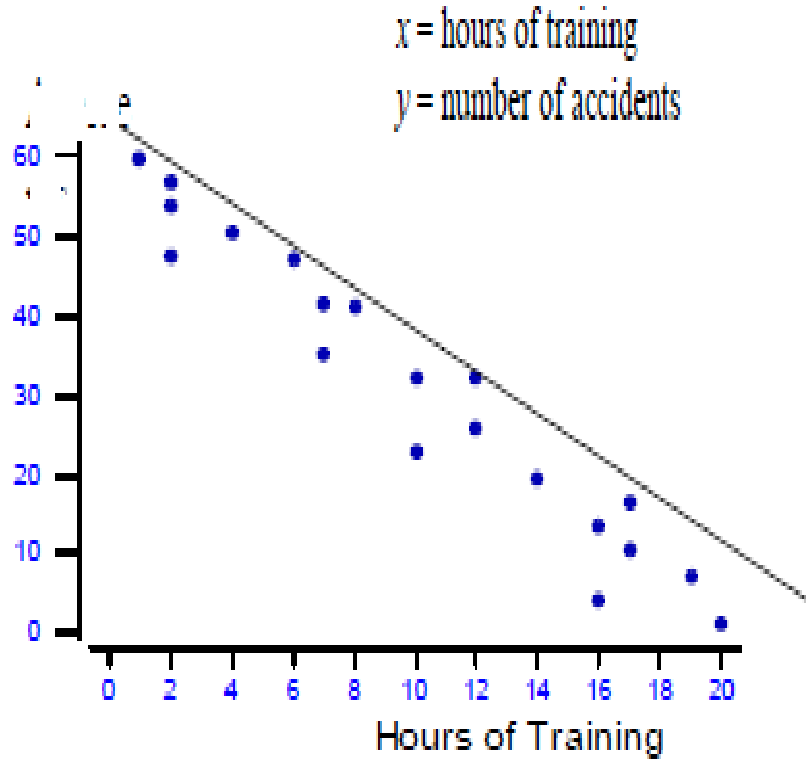


2	2 8 9
3	4 6
4	1 3 7 9
5	1
6	1 6
7	2 7
8	2

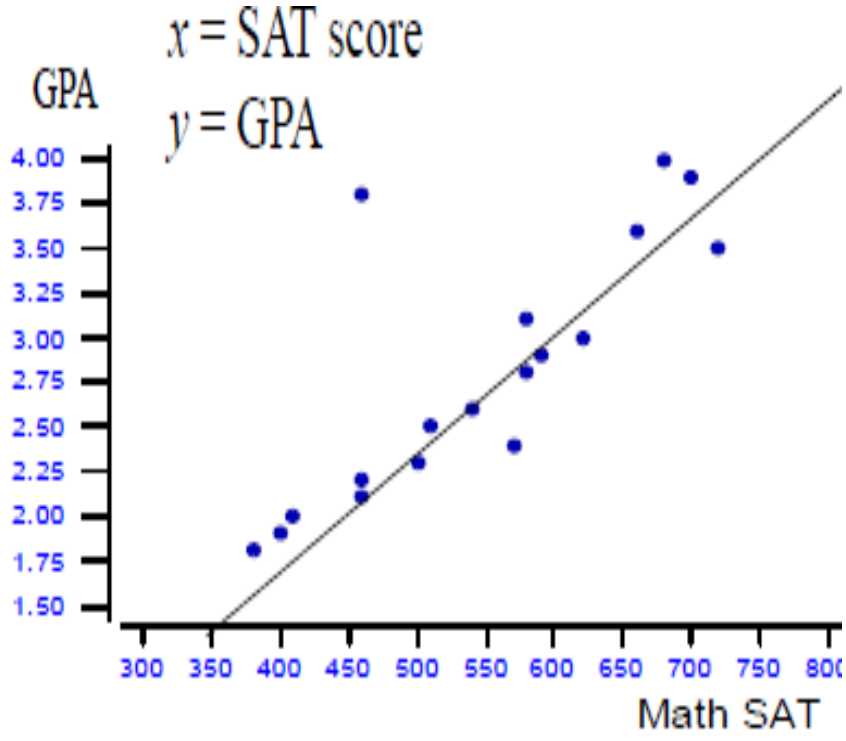
A box and Whisker diagram



Scatter plot



Negative Correlation as x increases, y decreases



Positive Correlation as x increases y increases

Thank you!!